DOCUMENT RESUME

ED 263 741 FL 015 189

AUTHOR Ingram, David E.; Wylie, Elaine
TITLE An Overview of Language Testing with Particular Reference to Indonesian in Australian Schools.
PUB DATE May 85
NOTE 25p.; Paper presented at the National Convention of the Indonesian Cultural and Educational Institute (1st, Melbourne, Australia, May 15-21, 1985).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Curriculum Development; Educational Objectives; Foreign Countries; Formative Evaluation; *Indonesian; *Language Proficiency; *Language Tests; Native Speakers; Second Language Learning; Summative Evaluation; *Test Construction; Testing; *Test Use
IDENTIFIERS *Australia

ABSTRACT

While a broad range of tests is available for second language testing, two of the basic approaches are developmental and non-developmental testing. The former is based on psycholinguistic evidence that language development is systematic, with all learners progressing through the same stages. The latter includes content selected in an ad hoc manner, on a linguistic or behavioral basis, with items and techniques chosen largely on the intuitions and experience of the examiner. The Australian Second Language Proficiency Ratings (ASLPR), a rating scale assessing general proficiency, is developmental in structure and is of the type that is most appropriate when the aim of the instruction is to develop practical second language proficiency. While the test, which can be adapted to any language, exists in a basic version for English as a second language and several other languages, development of a version in Indonesian would assist in testing learners of Indonesian as a second language. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED263741

AN OVERIVEW OF LANGUAGE TESTING WITH PARTICULAR REFERENCE TO INDONESIAN IN AUSTRALIAN SCHOOLS

D. E. INGRAM
Department of Education Studies
Darwin Institute of Technology

and

ELAINE WYLIE
Kingston High School, Brisbane

Invited paper to the First National Convention of the Indonesian Cultural and Educational Institute, Melbourne 15-21 May, 1985

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.E. INGRAM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 015189

1. INTRODUCTION

In this paper, we were asked to provide an overview of language testing and to relate it particularly to the teaching and testing of Indonesian in Australian Schools. Consequently, the paper will look briefly at the range of options available in second language testing and consider the appropriateness of these testing options for different testing purposes, it will seek to demonstrate how informed approaches to testing can have beneficial effects on language syllabus design, and it will seek to suggest solutions to some of the problems referred to, especially the problem of coping with native, non-native and partial native speakers of the target language in the same foreign language classes.

II LANGUAGE TESTS: THEIR NATURE AND PURPOSE:

In this paper, "assessment" or "testing" refers to the considered judgement that teachers make about student learning or student performance and "test" refers to any activity used in evaluating or measuring some part or all of a learner's language proficiency or other performance [cf. Ingram forthcoming]. There are many types of tests, which are not in competition but which, rather, serve different purposes, with the choice of test type depending to some extent on the physical and time constraints but, most importantly, on the aim of the assessment or testing. Thus, if one's interest is in how well the learner has formally learned the actual content presented during the course, then test based specifically on that content, i.e., on the syntax, vocabulary or functions specified in the course content, will be most appropriate. If the aim of the testing is to measure progress in a "graded objectives" or "graded levels of achievement" course, then the test should focus on the situations, tasks, functions, and their exponents specified in the course content. If, however, the aim is to assess the learner's general proficiency (or "competency" as it is called in recent syllabuses in Queensland), then a test that gives a picture of the learner's practical language skills in general, everyday situations will be most appropriate. Proficiency rating scales such as the Australian Second Language Proficiency Ratings (ASLPR) have been developed and seem most suitable for this last purpose. We will consider the ASLPR further later in this paper.

There are two broad approaches to test development that can be identified: the non-developmental approach and the developmental approach.

Non-developmental tests select the content to be tested in an ad hoc manner, on a linguistic basis, or on a behavioural basis. Ad hoc tests use items and techniques chosen largely on the intuitions and experience of the examiner. Many discrete point tests such as CELT and TOEFL as well as most traditional examinations fall into this category. Non-developmental tests prepared on a linguistic basis set out from a concept of the nature of language or the relevant domain of

the language which they seek to sample, or within which they seek to assess the learner's ability to operate; some use frequency counts of lexical or structural items to facilitate sampling, others sample randomly (e.g., cloze or dictation), while others use contrastive linguistics to identify "difficulties" or points of interference. Behaviourally-based non-developmental tests identify a particular language behaviour or aspect of language behaviour to be tested. This might include particular functions, tasks in specified situations (as in many "graded objectives" aproaches), or the language skills needed for some activity (e.g., an academic course).

Developmentally-based tests derive their justification from the psycholinguistic evidence that language development (whether first or second language) is not random but systematic, with all learners seeming to progress through the same stages. Thus, developmentally - based tests seek to relate a learner's ability to the stage of development he or she has reached in the developmental schedule. In the ASLPR and other scales which rate proficiency (defined as the ability to carry out language tasks), the levels are defined by describing the sorts of language tasks learners can carry out and how they are carried out in terms of such features as sentence forms, vocabulary range, formal accuracy, pronunciation, fluency, comprehensibility, rate of utterance, register flexibility, and so on. In using rating scales to measure proficiency, one asks what tasks the learners can carry out, how they carry them out, what is the nature of the language behaviour observed as the learners try to communicate, or, essentially, what behavioural description on the scale best matches the observed behaviour. Whereas rating scales designed to measure general proficiency describe the tasks and how they are carried out in global terms (in terms of an overall picture of the language behaviour observable at each level as learners attempt to use the language to carry out the tasks), in the graded objectives or graded levels of achievement approach, the actual tasks and the actual exponents by which they are carried out are generally specified.

We have noted that rating scales such as the ASLPR are developmental in structure, i.e., the progression of behavioural descriptions attempts to reflect the way in which a second language develops from zero to native-like. Such scales are not arbitrary but are related to the universal developmental schedule, and a learner's rating indicates the stage of development he or she has reached in the developmental schedule. This is important for at least three reasons. First, the proficiency measurements provided are not arbitrary in the sense that a percentage or score such as 4 on a 7-point scale or 495 out of, 550 or 600 is arbitrary and meaningless, but, in that they are behavioural and task-oriented, they tell the user exactly what the learner is capable of doing, and, in that they relate to the developmental schedule, the behaviour specified exists within a progressive, developmental sequence of behaviours. Second, although there are changes in all aspects of language behaviour (syntax, vocabulary, discourse, functions, and so on) throughout the developmental schedule, nevertheless different aspects and, especially, different parameters of change become salient at different developmental stages. These key parameters help the rater to identify

where the learner's proficiency falls on the scale. The parameters include, at the lowest levels, the range of situations in which the language can be used and a change from formulaic to more creative language, (0+ to 1 on the ASLPR), satisfaction of survival needs (1 to 1+), emerging complexification that facilitates the creation of social relationships (1+ onwards), register sensitivity and flexibility (3), and ability to see beyond surface meaning and to draw on significant cultural understanding (4 to 5). Third, because rating scales such as the ASLPR are behavioural and developmental, they indicate to the syllabus writer what learners at different levels can do, what their future directions of development are, and, therefore, what is most relevant to include in their course in order to maximize their rate of development.

These two broad approaches to developing tests, the non-developmental and developmental, lead to different types of test, commonly categorized into indirect, semi-direct, and direct tests. Indirect tests essentially test one thing, e.g., syntactic knowledge, and try to say something about something else, generally proficiency. This is done, usually, by psychometric, norm-referencing procedures in which the results, distributed desirably over a normal curve, are categorized in some way into proficiency levels. Typical of indirect tests are discrete-point tests in which language knowledge and language behaviour are analysed into the smallest possible units and knowledge of or ability to use those units is assessed. One major problem with this approach is in its interpretation: a score of 4 out of 7, 80%, or 495 out of 550 says nothing about the level of the learners' practical skills, about what they can do. Furthermore, discrete-point tests are analytic and isolate language items, whereas, in real-life, language items operate together, supporting each other in meaning, and dependent on each other structurally; part of the skill of language use involves being able to put all the items together and to comprehend them when received together. Semi-direct tests (also called "integrative" tests), such as cloze or dictation, do seek to integrate the language components into a total language event and to test the learner's knowledge of them and ability to use them in that total event. Though such tests resemble indirect tests in that they sometimes seem to be testing discrete items and the scores are processed psychometrically similarly to the scores on indirect tests, nevertheless, the fact that a total language event is used puts these tests closer to real language performance or the demonstration of real language proficiency, and hence the term "semi-direct tests". Direct tests (which are also "integrative") focus on actual language behaviour, i.e., they focus on the learner's proficiency as demonstrated in the way he or she carries out actual communication tasks. They assess directly what the testers are interested in. Statements about proficiency are made as the learners' language behaviour is matched against the level on a developmental scale consisting of a series of proficiency descriptions which best describes their language behaviour. Since the test itself and the statement of the learner's proficiency focus directly on what the learner can do, interpretation of the test is much easier and the result more immediately and practically useful.

III COMMON PURPOSES FOR TESTING AND SUITABILITY OF TESTS:

Earlier it was stated that different test types are not in competition but rather that they serve different purposes. It is useful, therefore, to consider some of the purposes for which one tests, some of the contentious issues in language testing, and what approaches to testing are most suitable to achieve those purposes.

III.1 A FUNDAMENTAL DISTINCTION: ASSESSING NATIVE AND NON-NATIVE SPEAKERS.

It is a fundamental principle in language testing, as in language teaching, that native and non-native speaker learners must be distinguished, i.e., one must distinguish the people who already know the language from those who do not. In other fields of education, such a statement would be considered trivial: if one were running tests in geometry, one would not give Euclid the same tests as a Year 8 mathematics student; one would not give Rutherford the same test as a Year 10 Physics student or vice versa. Yet, it remains the case that many school systems give the same test in a language to the Euclids and Rutherfords of a language as they do to early learners of the language. Then they wonder why the non-native learners are unable to compete, why the native speakers are not differentiated, or, perhaps worse, why it is necessary to emphasize artificial rules and language learning conventions unrelated to real communication skills in order to let the non-native speakers compete and to make some native speakers fail. It is absolutely fundamental that native and non-native speakers have to be distinguished, as is demonstrated by at least the following arguements:

1. There are radically different assumptions underlying the teaching and testing of native and non-native speakers. In particular, native speakers from a very early age have already internalized a fully developed linguistic system, while non-native speakers are still developing their understanding through the course being taught and tested.

2. Consequently, the purposes for teaching and testing native and non-native speakers are (or should be) different: for non-native speakers (i.e. L2 or FL learners), the primary aim of teaching is to develop the linguistic system, and of testing to measure where in the developmental schedule the learners are at and whether, or to what extent, they are able to communicate; for native speakers (i.e., L1 learners), the aim of teaching is to ensure that socially approved linguistic conventions that may be largely irrelevant to basic communication are adopted, and to develop such educated skills as different writing styles, appreciation of literature, and so on. For example, for a student to reach "Very High Achievement" in Senior English in Queensland, all or almost all of the students' work must (inter alia)

. demonstrate a high degree of sensitivity to the relationship

between audience, purpose and form in the selection of language;
- use language and the current conventions of writing and speaking with a high degree of accuracy;
- show clear understanding of the materials studied and be able to form and express personal opinions of sophistication and maturity.

Examples of material studied in a typical school work program are
- Tess of the D'Urbervilles
- Macbeth
- One day in the Life of Ivan Denisovitch.

Examples of tasks used in assessment are
- critical or creative essays (600-650 words)
- Seminar presentation
- Library research.

If this fundamental distinction between native and non-native speaking learners is not made, native speakers already having a more or less complete linguistic system will find the course taught of no interest and their test results will cluster at the top of the scale. At the same time, non-native speakers, especially in a normative system, will seem to perform so unfavourably that they will become discouraged, and the test results will be meaningless as indicators of their ability.

3. All language teaching is culture teaching; all language testing makes assumptions about the cultural knowledge of the learners. Native speakers may not necessarily be experts in their "high" culture but they have vast knowledge of the general culture or meaning - system underlying their language and acquired in the course of acquiring the language in interaction with other native speakers in the context of the language's culture. Non-native learners have not only a partial linguistic system but also a partial cultural knowledge. Consequently, the linguistic performance of native and non-native speakers is different [see Elias and Ingram 1977], the focus of language courses has to be different, and the appropriate tests must be different.

4. Empirical studies amply demonstrate the critical nature of this fundamental distinction between native and non-native speakers and also that a test designed for the former cannot be used successfully with the latter, and vice versa. In fact, it is a desirable strategy in validating a test for L2 learners to run the test with native speakers in the expectation that, if the test is a valid measure of where the learners are in the developmental schedule, the native speakers will cluster towards the top. In trialling some of the versions of the ASLPR, native speakers were included and were all rated at 5 (native-like). Again, Carroll in reviewing the CELT test states:

> "The authors are to be commended for presenting data on the performance of native English speakers on the tests - data that shows that American high school students of only moderate ability achieved 'extremely high scores' on them. Thus, one may infer that the tests measure fundamental rather than high-level skills in English."
> [John B. Carroll in Buros 1975, Section 7:260, p. 206]

Similarly, Angoff and Sharon [1971] showed that, on the "Test of English as a Foreign Language", native speakers were not differentiated and clustered at the upper end of the results.

5. It has been amply demonstrated in Victoria especially by Kelabora [e.g., 1983] that the indiscriminate mixing of native and non-native speakers does not work: if the tests are intended for FL learners, the native speakers all end up towards the top of the scale, while non-native speakers (especially in a normative system) fall at the lower end and are disadvantaged. As Kelabora has shown [1983], the inevitable result of this in a competitive and norm-referenced matriculation system is that non-native speakers will realize they cannot compete, that they are disadvantaged in taking the language, and their enrolments will decline. On the other hand, if the course is aimed at teaching and testing what they already know (viz., the linguistic system of the language), native speakers will lack stimulation and development, they will take the language as an easy option, an easy road to matriculation, and their educational and cultural development will be less than it could have been with a more appropriate course.

There would seem to be two broad solutions to this basic need to differentiate native and non-native speaking learners:

1. Taking account of the student population and assuming that it is a FL or L2 programme, one might identify goals and consequent course content and assess against these. Desirably, if the course aims at developing language proficiency, criterion-referenced assessment will be adopted using proficiency rating scales or a graded objectives approach. Anyone who meets or surpasses the criteria will be assessed accordingly, native speakers will cluster at the top of the scale and gain little from the course, while the non-native learners, if the numbers are large, will tend to distribute themselves throughout the scale. One might rationalize the lack of stimulation to the native speakers and the unbalanced distribution by shrugging one's shoulders and saying that, if native speakers are advantaged here, that is only a minor compensation for the disadvantage in other areas of the curriculum where they have to operate in English.

2. In reality, we have heard enough argument and illustration to show that native and non-native speakers must be differentiated both in their courses and in their assessment. In fact, where the student population is large enough, it is probably desirable in some languages (though possibly less so in Indonesian that in Greek or Italian in Australian schools) to stream learners into native, partial and non-native groups. The partial group, which is almost infinitely sub-divisible, will include those with a smattering of the language from home and, in some cases, those who speak a different dialect of the language.

The essential principle of both course design and assessment is pertinent: identify the purpose of the course and the assessment and

the most appropriate content and instruments by which to achieve those purposes. The purposes for native speakers and non-native speakers in studying a language are different, their starting-points are different, their courses must be different, and their assessment must be different. At the same time, the teacher and course designer should take advantage of the presence of native speakers to enhance the communicative activities that are possible in the classroom room, in particular by using community involvement approaches as outlined in other papers [e.g. Ingram 1978, 1979].

The following sections of this paper will, for the most part, assume that our interest is in the teaching and testing of non-native learners (i.e., FL or L2 learners) and that L1 learners are treated separately, and we shall consider what test types are relevant to achieve some common purposes in second language testing.

III.2. ELUCIDATING SYLLABUS DESIGN AND PROGRAMME PLANNING:

Indirect, especially discrete-point, tests can provide specific information on the items of knowledge or skill that contribute to or are missing from the learner's interlanguage. Thus, for example, a test such as the following shows whether the learner knows and can apply the rule that you use the conjunction than after a comparative:

Example 1: Tick the item that best fills the gap.

Question: "How old is Joe?"

"He's two years older _____ his sister Mary."

Answer: ☐ that ☐ as ☐ than ☐ of ☐

The test will not, however, clearly distinguish whether this rule is in the learner's conscious, learned knowledge and so only available for use when he has time to monitor or whether it is in his acquired knowledge and available for use in real-life situations. This, incidentally, accounts for the fact that indirect tests correlate at a relatively low and unstable level with direct tests of practical proficiency when the former are used with foreign language learners who, unlike second language learners, have had little opportunity to apply their knowledge in real-life, practical communication [see Ingram, forthcoming, for a fuller discussion of this point]. This observation is especially important if one is concerned with end-of-course assessment where the FL course has aimed at developing proficiency (e.g. in school courses) since it suggests that the approach usually adopted in matriculation examinations will not be valid or reliable. From a syllabus design or programme planning point of view tests of this sort most effectively provide diagnostic information on the specific items of syntax, pronunciation, vocabulary,

discourse, or functions that the learner does or does not know and which, if not known, the teacher may wish to include at some point in the course. In that such tests are based on specific items within the multitude of items that make up a language, there is a problem of sampling to ensure the syllabus designer or programme planner obtains adequate information.

Semi-direct tests are sometimes claimed to go part of the way towards solving this sampling problem because, if properly constructed, they are said to randomly sample the language or the domain of interest and they test the learner's ability to handle the items integratively within normal language functioning. For programme planning purposes, random cloze could well be supported by selective cloze focussing on particular features. Cloze is particularly useful as a means of identifying how the learner processes language, e.g., on a word-by-word basis, using an immediate context, or using the total context, and so it can effectively diagnose the learner's facility in using the cohesive and discourse features, so indicating the relevance of these features and the related linguistic processing for the syllabus or programme being developed.

"White noise" and interlinear tests can be used similarly to cloze, with the "white noise" test allowing one to focus on ability in listening. Dictation may be used similarly if presented as Oller proposes [1979: Chap. 10] so that the dictated chunks exceed immediate memory capacity and introduce a productive or creative element in the test. Delivered more slowly and in smaller chunks, it may be used to identify phonological needs and, in writing, sound-symbol correspondence or spelling needs. Brindley and Singh [1982] of the N.S.W. Adult Migrant Education Service base their testing procedures on the assumption that a syntactically detailed universal development schedule has been identified and use dictation incorporating specific syntactic items to locate the learner on the developmental schedule. This approach provides information on syntactic needs in programme planning as well as, it is assumed, on the other areas of development that go on in conjunction with syntactic development.

Direct testing with developmental instruments such as the ASLPR may contribute in at least two ways to syllabus design or programme development. First, as noted earlier, in that certain major parameters of change emerge at different stages in the scale, the learners' location at any particular point on the scale indicates the development trends that are dominant for them and that a course should attempt to encourage. Second, such direct testing may also be used to identify specific times of need in the learner's language. More precisely, it is careful observation of the leaner, possibly with the use of a checklist during the interview procedure by which the tests are applied that is used. The major problem with this approach to identifying developmental needs and so desirable course content is that it depends to some extent on chance or, at best, on the skill of the interviewer whether or not the comprehensive needs of a learner will be elicited. In addition, it is a time-consuming process and it is difficult to interview and at the same time maintain notes or even a

checklist of learner needs. Nevertheless, a skilled interviewer can obtain much information about a learner's language needs.

III.3. SETTING GOALS AND OBJECTIVES:

In syllabus design and programme planning, it is desirable to distinguish goals and objectives and different tests make different contributions to them. Goals are long-term or overall aims identified in terms of the nature (e.g., general or "special purpose") and level of the proficiency sought. As one stands off from the teaching programme, the goals are the points to which the teaching and learning activities are ultimately leading the learners. In expressing goals, there is no place for such vague "motherhood" statements as "to develop the ability to converse freely with native speakers". Rather, a direct instrument such as the ASLPR lets goals be stated specifically in terms of the exit behaviour sought, i.e. for example, in course x, the goal may be to develop in the learners the following proficiency level: S:3, L:3+, R:2+, W:2, which, in turn, is translatable (especially in graded objectives courses) into the range of situations in which the learners will minimally be able to use the language and the linguistic forms that will occur. Once the learners' starting-point and the goals are specified in this way, then, as already noted, the ASLPR also reveals the dominant developmental trends to be catered for in the programme.

Objectives are the on-going aims the teacher adopts as he or she selects language content to teach and activities for the learners to participate in. The objectives are short-term aims, especially immediate formal learning aims, which the teacher has in mind in presenting material. Language content can be identified both using the needs-based approach to syllabus design and programme planning discussed later and by diagnosing the learner's linguistic deficiencies using indirect, semi-direct and direct tests as discussed earlier. Nevertheless, if viewed by an impartial observer, the items identified for teaching in this way (unless, perhaps, they were precisely identified using a developmentally-based discrete point test) would seem to be not so much components of language behaviour that are being developed as items of language experience the teacher is making available to the learners. Thus, for instance, the needs analysis or the test used may lead the teacher to teach the present continuous tense with performance objectives such that, by the end of the lesson, the learners will be able to say such utterances as "He is closing the door" or "I am reading a book". However, Krashen"s distinction between learning and acquisition suggest that, unless the learner is "ready" to acquire that verb form, he will not use it in a natural situation outside the formal classroom (except where he can consciously "monitor" his language) [see, for example, Krashen 1976]. The teacher, in other words, may quite correctly have identified the present continuous and its formal learning as an objective but its teaching has, in reality, been a means of providing the learner with systematic language experience that will help his natural learning strategies move him towards the course's specified proficiency goal. In brief, identifying goals and objectives assists the teacher to judge at all times whether

the language content being taught and the activities chosen are contributing coherently to the learner's on-going language development.

III.4 TESTS AS SYLLABUSES:

At no time should tests be considered to be syllabuses (though, regrettably, external examinations have traditionally assumed this role): the needs that produce an instrument to measure proficiency are not the same as the factors that determine how a language is learned, nor is the goal to be attained, viz. a specified proficiency level, the same as the developmental path followed in moving towards that goal. While the ASLPR, for example, can appropriately be used to specify proficiency at different levels, none of those descriptions should be seen as constituting a syllabus, i.e. the examples of tasks that can occur at S: 2, L: 2, R: 2, W: 2 can not be practised or rote memorized as a means of providing the learner with that level of proficiency. First, they are only examples of tasks and do not constitute a complete, all-inclusive picture of what can be done at that level; second, as already noted, a goal is different from the path towards it; and, third, the creativity that underlies proficiency at all but the lowest levels cannot come from rote-memorizing task examples of a proficiency level. Nevertheless, a scale such as the ASLPR does indicate the learner's stage of development, what activities are relevant to his or her stage of development, and what path this future development will take. This has clear implications for syllabus design as already discussed but it does not specify detailed content.

III.5 FORMATIVE AND SUMMATIVE ASSESSMENT:

The contrast between goals and objectives discussed earlier implies that there will be a difference between assessment at the end of a course, especially assessment of the general proficiency attained, and on-going course achievement, i.e., between summative and formative assessment. If the goals are specified in terms of general proficiency, then direct instruments such as the ASLPR that measure general proficiency are most appropriate, and should be used at exit-points (i.e., for summative purposes). It may also be that, in considering course attainment, one wants to know to what extent the learner has internalized the content of the course and, if so, then tests directly related to that content may be most appropriate. For example, if one wants to know whether the specified grammatical content has been mastered, then discrete point tests of the grammar taught may give the most direct information. A functional or language task checklist as used on the Lothian progress cards and Waystage Tests may be used to assess achievement in the skills aspects of the course [see Clark 1982 and 1982a].

At intermediate points in a course, it is probably not appropriate to use the ASLPR or another test of general proficiency because general proficiency changes too slowly (especially in unintensive Secondary School foreign language programmes) for valid and reliable

measurements of general proficiency to show much change or for useful information to be fed into on-going course development. At intermediate points, teachers are probably most interested in formative testing, in particular in order to assess what of their formal teaching is being learned and, possibly, acquired. What test will be most appropriate for the purpose will depend on what the teacher wants to find out, especially in order to feed back into the programme planning. If the teacher is interested in whether or not the learner is internalizing the grammar being taught, then test items that focus specifically on the grammatical rules will be most appropriate. If the teacher wants to know if the student can perform certain functions, then a role play approach or even an item such as "Tell me what you would say in a post office to ask the cost of sending a letter to Australia" might be used. To be valid, such test items must, of course, be contextualized in a realistic situation with a contextual purpose for the transaction to occur. The Lothian progress cards and Waystage Tests provide well-structured ways to check the functional and activity content of the course [see Clark 1982 and 1982a]. The general principle to testing is applicable to formative and summative assessment: find out what information is needed and select the test and test type accordingly.

III.6 CONTINUOUS V's ONE-OFF EXIT-POINT ASSESSMENT:

The issue of summative and formative assessment is relevant also to the question of continuous or one-off exit-point assessment. It has become fashionable in education to speak favourably of continuous assessment and to deplore one-off, end-of-course or exit-point assessment. However, such criticism of exit-point assessment ignores the nature of skill development in general as well as the distinctive nature of language learning. First, if a learner is to commence another course, start work, go overseas, make some other use of the language or simply leave the programme, it is his or her proficiency at that point that is of interest and not what it was 6, 12 or 24 months earlier. It may be that the teacher's continual monitoring of each learner's performance will facilitate the final assessment, and checklists of what the teacher observes the learner can do may be worthwhile, but an exit-poing test using the interview procedures recommended for use with rating scales is essential.

Second, we noted earlier that there is an obvious and important distinction between a long-term goal (such as attainment of a particular proficiency level) and the path followed in attaining that goal and, as Krashen noted [e.g. 1976], between "learning" and "acquisition", between what is formally taught, learned or memorized and what is able to be used in practical language tasks reflecting the creativity and spontaneity of real language use and general proficiency. Briefly, this means that a student may be able to formally learn aspects of the language (e.g., through rote memorization and the repetitive application of rules in formal exercises) and yet be unable to use the rules or vocabulary when it comes to real-life communication tasks. If we are interested in measuring practical language skills or practical proficiency, it is the learners'

proficiency as reflected in real-life communication tasks that is important and not their formal knowledge. Continous assessment is more likely to focus on the latter (i.e., on their formal learning of the course content most recently taught) and is unlikely to provide a true picture of what has been acquired and is available for use spontaneously and creatively in practical language tasks.

For these reasons, rather than basing proficiency or competency assessment on continuous testing, it would seem desirable for teachers to place most emphasis on exit-point assessment. What then is critical is that this summative assessment be so conducted that the learners' maximum language behaviour is elicited, observed and rated. This means that, for instance, interviews and other tests must be in accord with recommended principles [cf. Ingram 1982: Section IV; forthcoming: Section III. 3.ii] and should be conducted, scheduled and organized so that each learner has an opportunity to demonstrate his or her real ability and is not disadvantaged by, for example, poor or unduly stressful procedures or such irrelevant issues as a day's illness.

Third, the contrast between continuous and exit-point assessment highlights again the fundamental distinction in testing between course attainment and general proficiency. Course attainment relates more to what learners have formally learned, to what they have taken from each of the teacher's lessons. Such testing can relate quite specifically to the content of the course and the test types could reflect quite closely the activities used in teaching. In fact, it is commonly "accepted wisdom" in education that assessment procedures should mirror the teaching and learning activities of the course. However, as demonstrated in the grammar-translation era of language teaching, tests of this sort, even though they may accurately reflect learners' attainments in the actual course taught, may give little information on the learners' real practical proficiency. Tests of course attainment may, in other words, be highly desirable for many purposes (e.g., feedback on the effectiveness of teaching strategies, success of formal teaching and learning, or for diagnostic purposes); and they will be used throughout a course. However, they should not be confused with measurements of general proficiency and, if the aim of the course is to develop general proficiency, the final assessment should be based around test types and actual instruments that measure general proficiency irrespective of the actual course followed. For this purpose, rating scales such as the ASLPR would seem most useful and should be used as the basis at least of the exit-point assessment.

III.7. ESTABLISHING SCHOOL GRADES:

One of the difficulties teachers sometimes have with rating scales is how to convert proficiency ratings to the sorts of grades traditionally used in Secondary Schools. Not least, if one is using the ASLPR, it may seem demotivating to state at the end of Year 10 that a learner has reached a proficiency of 1- (a realistic level for Indonesian) in all four macroskills. Such concern overlooks several important issues. First, there is no doubt that much of the

disillusionment students feel with language study comes from false and unrealistic expectations. That is, on the one hand, they seem initially to believe that they will rapidly become like native speakers and they become disillusioned when they realize that language learning is a long and often arduous process and, on the other hand, they believe (partly as a result of the exaggerated importance schools have traditionally attached to detailed formal accuracy) that they cannot communicate until they know and use the language with the accuracy (or, even, better than the accuracy) of native speakers.

The response to this disillusionment should take two related forms. First, if teaching at all times emphasizes communication, not only will it ensure that real, purposeful and meaning-oriented language and realistic language use are being taught, but the learner will also realize at all times that he can and does communicate despite fragmentary utterances and a limited range of vocabulary and other features. In addition, there is invoked the same basic motivation (viz., the desire to communicate more adequately) as occurs in natural first or second language learning. Second, syllabus writers and teachers should set clear and realistic goals and should avoid seeming to promise the dishonest, unrealistic attainments that language syllabuses have traditionally implied. The Queensland German syllabus, for example, sets as the proficiency in Speaking for "Very High Achievement" in Year 10 as:

> Student demonstrates ability to ask and answer simple questions on the topics prescribed without hesitation. He can maintain a simple face-to-face conversation. He can speak with no significant errors in pronunciation, intonation and stress. Errors in grammar, syntax and vocabulary are few, and of a minor nature, and should not prevent comprehension by the listener. He shows a mastery of simple idiom and can use a variety of vocabulary within the prescribed topics. Self-confidence, rather than self-consciousness is evident.

Such an expectation of non-native speakers in, at best, a 220 hour course spread over three years is entirely unrealistic and creates false expectations. Kelabora [1983] has noted that the Victoria Indonesian syllabus is equally unrealistic. Rating scales provide a ready means to set goals quite precisely. Thus, as we have seen, one might state that the goal in a three year unintensive French course (Years 8 to 10) is the proficiency S:1-, L:1-, W:1-, R:1- or, in Japanese, S:1-, L:1-, W;0+, R:0+. Probably, for Indonesian, the same levels as for French would be appropriate though there is some evidence [FSI 1973] that they may be a little high.

Furthermore, this approach to specifying goals readily lends itself to conversion into whatever grades (e.g., words, numbers, or letters) an education system chooses to require (though a profile is probably more meaningful). Thus, if goals are specified in terms of the proficiency levels the average student could be expected to achieve, then achievement above or below this leads to higher or lower

grades. The present writers' own recommendations, for example, to the Foreign Languages Advisory Committee of the Queensland Board of Secondary School Studies on the specification of goals for the unintensive Queensland Secondary School French courses and the conversion of proficiency levels attained to Board grades were as follows:

YEAR 10					
GRADES	PROFICIENCY				
	S	L	R	W	OVERALL
Very High Achievement	>1	>1	>1	>1	All four macroskills >1
High Achievement	1	1	1	1	Two or more macroskills 1, others 1-
Sound Achievement	1-	1-	1-	1-	All four macroskills 1-
Very Low Achievement	0+	0+	0+	0+	1 or more macroskills 0+
Low Achievement	<0+	<0+	<0+	<0+	Two or more macroskills <0+
					NOTE: Any variation below the number of macroskills at a specified level of achievement would have to be compensated for by an equal number above for the learner to be graded at that level of achievement.

YEAR 12					
GRADES	PROFICIENCY				
	S	L	R	W	OVERALL
Very High Achievement	>2	>2	>2	>2	All four macroskills >2
High Achievement	2	2	2	2	Two or more macroskills 2, others 1+
Sound Achievement	1+	1+	1+	1+	All four macroskills 1+
Low Achievement	1	1	1	1	One or more macroskill 1
Very Low Achievement	<1	<1	<1	<1	Two or more macroskills <1
					NOTE: Any variation below the number of macroskills at a specified level of achievement would have to be compensated for by an equal number above for the learner to be graded at that level of achievement.

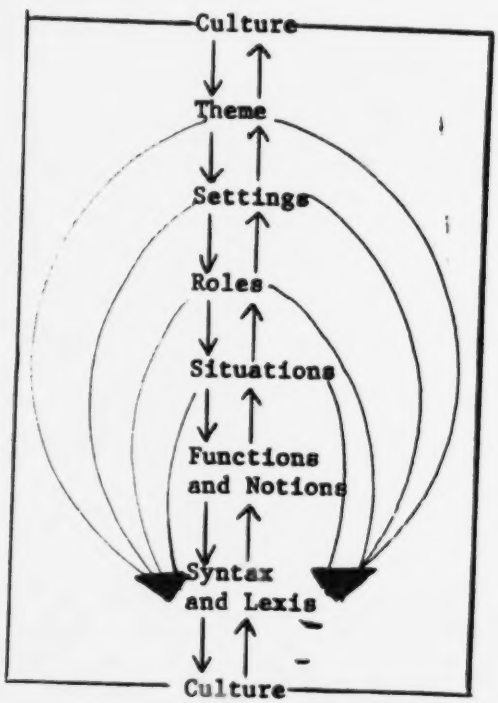
IV BASIC PRINCIPLES IN SECOND LANGUAGE SYLLABUS DESIGN

Because of the concern that has been expressed about the nature of Indonesian syllabuses [e.g., Kelabora 1983], it seemed appropriate to outline some basic principles in modern approaches to syllabus design. A language programme is, or course, about much more than the development of proficiency (effecting attitude changes, for example [see Ingram 1978, 1979] but here we shall focus only on the language component.

The needs in the linguistic component of the syllabus design task include how to identify goals, how to indentify what is relevant to the learners' stage of development, how to ientify course content, and how to ensure that the course is coherent and systematically develops the learners' language skills rather than inundate them with developmentally irrelevant language or a multitude of discrete items to be rote-memorized. Earlier we noted the role of language tests (especially direct instruments such as the ASLPR) in setting goals and objectives, in identifying major developmental trends and relevant activities at different stages, and in providing an overarching framework within which a coherent syllabus can be planned. Here let us focus principally on the specification of language content in response to learner needs. These principles are not new, they have become well-established in the literature in the last ten years and, like most of the rest of this paper, we have written or spoken on them in many places.

Language occurs in situations, it is used in situations, and a learner's realization of need comes from experience in situations. Consequently, the starting-point in planning a course and relating goals to language content is to identify the <u>situations</u> in which the language is to be used (see Table 1 and Fig. 1). More precisely, one must first ask in what contexts or <u>settings</u> the learner needs to be able to use the language. So, for example, a new arrival in a country (whether a migrant, a tourist or a student) may need to use the language in a shop, in a doctor's surgery, in a post office, at a bus stop, and so on. A science graduate may need to use the language in a science laboratory, in a library, or in the staff common-room. When a setting has been identified, then one can identify the <u>role</u> the learner will fill in that setting. So, for example, in a shop, a person may fill the role of a customer or a shop assistant; in a doctor's surgery, the role could be that of a patient, the interviewing doctor or a receptionist; and so on. When the settings and the likely roles in them have been identified, then one can identify the purposes for which the language will be used in the settings and so the language transactions or language <u>events</u> that will take place and, therefore, the language <u>functions</u> that the learner will need to be able to carry out and the ideas or <u>notions</u> that he needs to be able to convey. The <u>language transaction</u> together with the <u>role</u> and <u>setting</u> constitutes a situation. So, for example, the learner may need to be able to use the language in the situation of enquiring where certain foods are located in a supermarket (setting: supermarket; role: customer; function: enquiring about location; notion: location).

Fig. 1 The Identification of Language Content



(N.B. The starting-point is with settings)

Table 1 Settings, Situations, Roles, Functions and Notions

SETTING	SITUATION	ROLE	FUNCTIONS	NOTIONS
Doctor's Surgery	Being interviewed by a doctor in his surgery	Patient	Describing symptoms, Seeking information about medication, etc.	Pain, Intensity, Frequency, etc.
Post Office	Buying stamps in a post office	Customer	Seeking information about cost, etc.	Price, etc.
Science laboratory in a College or University	Writing up an experiment in a science laboratory	Teaching Assistant	Describing substances, Narrating chemical events etc.	Colour, Texture, etc.
Staff common-room in a College or University	Conversing in a staff common-room	Staff Member	Greetings, Entering a conversation, etc.	Courtesy, etc.

(Note that, if a table such as this is to be used as a basis for course design, the following additional columns should be added: Syntax, Lexis, Non-Verbals, Other Linguistic Features, Cultural Knowledge).

When the situations, functions and notions have been identified, then the syntax and vocabulary or lexis needed to carry them out can also be identified. So, for example, the function of describing symptoms will require such syntax as NP + V + Adv. Time, possessive pronouns, and simple present tense and it requires such lexis as parts of the body, "hurt" and time expressions (e.g., "My arm hurts all the time"). Seeking information at a bus stop about direction of travel requires such syntax as do-questions and lexis such as "go", "past" and names of streets and places (e.g., "Do you go past the post office in Bourke Street?").

When one identifies all the settings in which a learner needs to be able to use the language, it is found that they tend to cluster into groups or themes. So, for example, settings such as a doctor's surgery, a dentist's waiting room, a hospital casualty department or an optometrist's studio, cluster into a theme of "health". This observation and the concept of themes are very important since, first, they provide a way of grouping and organizing the multi-various settings in which a learner may need to be able to use the language and, of a least equal importance, themes enable one to identify and organize the minimum cultural information a learner needs if he is to be able to operate in the language. Thus, for example, the theme "health" in the settings just listed suggests such essential cultural information as male-female and doctor-patient relationships, how the health services are organized, and what health insurance provisions exist; the theme of "transport" arising from settings such as "at a street corner", "at a bus stop", "at a railway ticket office", and "at an airport check-in desk" suggests cultural information such as how crossing a city street is regularize, what side of the road cars drive on, private and public transport arrangements, road safety provisions, and traffic control. The identification of cultural information in this way is important for several reasons. First, meaning conveyed by language exists only in the context of the culture and, deprived of the culture that is integrated with it, language reduces to a sort of meaningless, verbal algebra; second, whereas language teaching traditionally neglected the real, everyday culture in favour of the atypical literary culture, this approach re-emphasizes the general culture; and, third, one needs a way of organizing cultural information and integrating it with the language teaching so that it can be taught systematically and coherently and not become a disjointed list of cultural snippets.

V CONCLUSION

In this paper we have attempted to provide an overview, albeit brief and necessarily cursory, of language testing. We have tried to draw attention to the range of test types available. We have emphasized that different tests serve different purposes, different purposes determined by different learner needs and course content demand different approaches to syllabus design and learner assessment. We emphasized, that direct test instruments such as the ASLPR are most appropriate where the aim of the course is to develop practical proficiency in a second or foreign language. The ASLPR has been

developed for use in assessing the proficiency of second or foreign language learners of any language. It now exists in a basic version using ESL examples, in versions for French, Italian, Japanese and Spanish as foreign languages, and three "English for special purposes" versions are currently being prepared. It is already available for use with any language but if an Indonesian version were available its use with ISL or IFL learners would be facilitated. We would be delighted to co-operate with anyone interested in working on an Indonesian version with us.

Underlying all of what has been said in this paper is the most fundamental principle of second language teaching: learning occurs in response to need, courses must identify and be designed to meet learner needs; the starting-point in the learners' own characteristics and the needs to be met by the course result in purposes, aims, goals and objectives which determine the nature and content of the course. The purposes, aims, goals and objectives of the course help identify the various purposes of testing and lead to the same basic and pervasive principle in language testing:

> Identify the purpose of the testing. Select the test type that most effectively achieves this purpose with the particular learners to be tested.

Alderson, J. Charles and Arthur Hughes (eds.). 1981. Issues in Language Testing, ELT Documents No. 111. London: The British Council.

Angoff, Willian H. and Amiel T. Sharon. 1971. "A Comparison of Scores Earned on the Test of English as a Foreign Language by Native American College Students and Foreing Applicants to U.S. Colleges". In TESOL Quarterly, Vol.5, No.2, June 1971: 129-136.

Australian Federation of Modern Language Teachers Associations. 1982. "Submission on Aspects of a National Language Policy", submission to the Standing Committee on Education and the Arts of the Australian Senate, November 1982. (mimeograph).

Brindley, Geoff and Ken Singh. 1982. "The Use of Seocond Language Learning Research in ESL Proficiency Assessment". In Australian Review of Applied Linguistics, Vol. 5, No. 1, July 1982 : 84-111.

Buros, Oscar K. 1975. Foreign Language Tests and Reviews. Highland Park, N.J.: Gryphon Press.

Clark, John L. 1982. "Graded Levels of Achievement in Foreign Language Learning (Lothian Region's Project in Modern Languages in Schools)". In Language Testing Newsletter, No. 2, Spring 1982 : 2-4.

Clark, John L. 1982a. "How are we to assess language learning in schools?" Paper to the Fourth National Language Conference, Perth, 3rd - 6th September, 1982. Reprinted in MacPherson 1982.

Elias, G.C. and D.E Ingram. 1977 Cultural Components or Reading. Singapore: University of Singapore/RELC.

Foreign Service Institute School of Lnaugage Studies (FSI). 1973. "Expected levels of Absolute Speaking Proficiency in Languages Taught at the Foreign Service Institute". Washington D.C. : U.S. Department of State. mimeograph.

Hyltenstam, Kenneth and Manfred Pieneman. 1985. Instructional and Social Implications of L2 Acquisition Research. London : Multilingual Matters, 1985.

Ingram, D.E. 1977. "Language Proficiency at Matriculation". In Australian Review of Applied Linguistics, Vol. 1, No. 1, December 1977 : 20-32.

Ingram, D.E. 1978. "An Applied Linguistic Study of Advanced Language Learning", thesis for the degree of Doctor of Philosophy, University of Essex, Cohchester, England. In ERIC Collection ED168359.

Ingram, D.E. 1979. "Methodology", base paper for the On-Arrival Section of the Adult Migrant Education Programme. Canberra : Department of Immigration and Ethnic Affairs. mimeograph.

Ingram, D.E. 1982 Introduction to the Australian Second Language Proficiency Ratings (ASLPR). Canberra : Department of Immigration and Ethnic Affaris/AGPS. 1979 rev. 1982.

Ingram, D.E. forthcoming. "Assessing Proficiency : An Overview on some Aspects of Testing". Forthcoming in Hyltenstam and Pienemann, 1985.

Kelabora, Lambert. 1983. "Bahasa Indonesia is going through a crisis in Victoria, Australia : Diagnosis of the Present Situation with Policy Options for the Future", paper to the Eighth Annual Congress of the Applied Linguistics Association of Australia, Melbourne, 29-31 August, 1983.

Krashen, S.D. 1976 "Formal and Informal Linguistic Environments in Language Acquisition and Language Learning". In TESOL Quarterly, Vol. 10, No, 2, June 1976 : 157-168.

McPherson, Anne. 1982. The Language Curriculum in the 1980's : Affective, Effective or Defective? Perth : MLTAWA/AFMLTA.

Oller, John W. 1979. Language Tests at School : a Pragmatic Approach. London : Longman.

Vollmer, Helmut J. 1981. "Why are we interested in 'General Language Proficiency'?" In Alderson and Hughes 1981 : 152-175.